



ФОРМАЛІЗАЦІЯ ІНФОРМАЦІЙНО-ЛІНГВІСТИЧНИХ СУТНОСТЕЙ ТА ВІДНОШЕНЬ ЗАСОБАМИ ШТУЧНОГО ІНТЕЛЕКТУ

Петрасова С.В.

*Національний технічний університет
«Харківський політехнічний інститут»,
м. Харків, вул. Пушкінська 79/2, тел. 707-63-60,
e-mail: svetapetrasova@gmail.com*

Однією із проблем сучасного суспільства є інформаційне переповнення, що вимагає переходу від зберігання та обробки даних до накопичення та обробки знань. У зв'язку з цим стає актуальним завдання інтелектуалізації лінгвістично-інформаційних технологій представлення та обробки знань.

Основним засобом накопичення та передачі знань при цьому служить природна мова. Однак з точки зору вирішення поставленого завдання природна мова має істотні недоліки: вона неповна, надлишкова, неоднозначна, неточна та граматично некоректна.

Неповнота природної мови виражається у відсутності в ній слів для вираження смислових особливостей (нюансів) в діалозі або нових понять.

Прикладом вираження надлишковості мови є наявність в ній слів-синонімів, що виражають один і той самий зміст.

Неоднозначність природної мови виражається в наявності в ній слів-омонімів, що мають різний зміст у різних контекстах.

Неточність природної мови виражається в «розмитості» шкали смислових інтерпретацій слів.

Граматична некоректність мови проявляється в наявності винятків у багатьох правилах мови [1].

Отже вирішення завдання автоматичної обробки знань ускладнюється проблемами полісемії, синонімії, омонімії природномовних джерел знань, які властиві мові на всіх рівнях її представлення (морфологічному, синтаксичному, семантичному і прагматичному), що, в першу чергу, проявляється в проблемі визначення відношення семантичної еквівалентності.

При цьому складність завдання підвищується у рази, коли мова йде про смислову близькість не слів як основної значущої, синтаксично самостійної одиниці мови, а багатослівних словосполучень (колокацій) – синтаксичних конструкцій, що складаються з двох або більше слів (колокатів), з'єднаних за типом підрядного зв'язку (узгодження, управління, примикання).

З наведених причин ідентифікація інформаційно-лінгвістичних сутностей, зокрема колокацій, та відношень в текстах з характерними для людського мовлення гнучкістю та багатозначністю вимагає інтелектуальних засобів семантичної обробки природномовних текстів.

Для розв'язання вказаних завдань застосовуються статистичні, морфологічні, синтаксичні та семантичні методи аналізу природномовних виразів, а також методи представлення формалізованих моделей зовнішнього світу, що є

предметом людино-машинного діалогу і сприяють розумінню машиною природної мови [2].

Тим не менш більшість існуючих формальних моделей смислової ідентифікації слів, рідше колокацій, не дозволяють автоматизувати процедуру ідентифікації або видають низьку точність при практичній реалізації. Це пов'язано з тим, що для реалізації автоматичного смислового аналізу природномовних текстів використовуються лексико-граматичні підходи, контекстний аналіз, статистичний аналіз, та рідше застосовуються семантичні методи у комплексі з іншими методами, що стає обов'язковим при ідентифікації інформаційно-лінгвістичних сутностей та відношень.

В дослідженні для розпізнавання відношення смислової близькості використовуються граматичний і семантичний описи таких мовних сутностей як колокації, що містять інформативну формальну складову.

Основна ідея підходу полягає в тому, що використання методів штучного інтелекту, а саме, апарату алгебри кінцевих предикатів, компараторної ідентифікації та компонентного аналізу, а також множини граматичних і семантичних характеристик для формалізації інформаційно-лінгвістичних одиниць дозволить локалізувати семантично близькі кореляції (відношення семантичної еквівалентності) в природномовних текстах.

Лінгвістично-інформаційна технологія покликана підвищити ефективність роботи існуючих інтелектуальних систем семантичної обробки текстів, наприклад, при усуненні смисловий неоднозначності, добуванні фактів та ін. за рахунок формалізації як синтаксичної, так і семантичної інформації та автоматизації виявлення відношення семантичної еквівалентності.

Таким чином, ідентифікація бінарних відношень семантичної еквівалентності між колокаціями, як найбільш конкретних змістовних інформаційно-лінгвістичних сутностей, дозволяє формалізувати відношення семантики, неявно виражене в природномовних конструкціях, та підвищити точність, повноту і релевантність запитуваної інформації.

Список літератури:

1. Иванов В.М. Интеллектуальные системы / В.М. Иванов. – Екатеринбург, 2015. – 92 с.
2. Шенк Р. Обработка концептуальной информации / Р. Шенк. – М. : Энергия, 1980. – 360 с.
3. Хайрова Н.Ф. Концептуальная схема идентификации смысла лингвистических единиц / Н.Ф. Хайрова // Збірник наукових праць Військового інституту Київського національного університету імені Тараса Шевченка. – 2013. – Вип. 39. – С. 217-223.
4. Бондаренко М.Ф. Мозгоподобные структуры / М.Ф. Бондаренко, Ю.П. Шабанов-Кушнарченко. – Под. ред. акад. НАН Украины И.В. Сергиенко. – К. : Наукова думка, 2011. – 460 с.